

Estimating Population Prevalence of Posttraumatic Stress Disorder: An Example Using the PTSD Checklist

Artin Terhakopian

Center for the Study of Traumatic Stress, Uniformed Services University of the Health Sciences, Bethesda, MD

Ninet Sinaii

Biostatistics and Clinical Epidemiology Service, National Institutes of Health Clinical Center, Bethesda, MD

Charles C. Engel

Center for the Study of Traumatic Stress, Uniformed Services University of the Health Sciences, Bethesda, MD

Paula P. Schnurr

National Center for PTSD, VA Medical Center, White River Junction, VT

Charles W. Hoge

Division of Psychiatry and Neuroscience, Walter Reed Army Institute of Research, U.S. Army Medical Research and Materiel Command, Silver Spring, MD

The PTSD Checklist (PCL) is among the most widely used self-report instruments for assessing PTSD. To determine PCL's performance on a population level, the authors combined data from published studies that compared the PCL with structured diagnostic interviews. Weighted average sensitivities and specificities were calculated for cutoff categories most often reported in the literature. Weighted average sensitivity decreased from .85 to .39 and specificity increased from .73 to .97 for cutoffs ranging from 30 to 60. The PCL's ability to accurately estimate PTSD prevalence varied as a function of cutoff and true PTSD prevalence. In populations with a true PTSD prevalence of 15% or less, cutoff values below 44 will substantially overestimate PTSD prevalence. Uncalibrated use of the PCL for prevalence estimation may lead to large errors.

Validated self-report measures are frequently used to identify persons who need clinical evaluation for a mental disorder or to measure symptom severity and treatment response. Self-report measures are also used to estimate the population prevalence of mental disorders or to screen for disorders such as posttraumatic stress disorder (PTSD) in populations exposed to traumatic events. Validation studies of self-report measures are virtually always based on clinical samples of patients from primary care or specialty mental health settings. The use of self-report measures for population prevalence or population screening is rarely supported with direct validation studies. Cutoffs are selected from studies conducted in clinical settings. This study estimates the population performance

of one self-report measure for PTSD, the PTSD Checklist (PCL; Weathers, Litz, Herman, Huska, & Keane, 1993). The PCL is one of the most widely used self-report instruments for measuring PTSD symptoms (Elhai, Gray, Kashdan, & Franklin, 2005).

Developed at the National Center for PTSD, the PCL is a 17-item self-report questionnaire based on the criteria given in the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)* (American Psychiatric Association, 1994). It takes less than 10 minutes to complete and requires a 10th-grade reading and comprehension level (Posttraumatic Stress Disorder Checklist-Military [PCL-M], 2005). Three versions of the PCL exist and are differentiated based on the identified trauma, e.g.,

No funding was sought for this project.

The views expressed in this article are those of the authors and do not reflect official policy or position of the Department of the Army, the Department of Defense, the Department of Veterans Affairs, the U.S. Government, or any of the institutional affiliations listed.

Correspondence concerning this article should be addressed to: Charles W. Hoge, Director, Division of Psychiatry and Neuroscience, Walter Reed Army Institute of Research, 503 Robert Grant Avenue, Silver Spring, MD 20910. E-mail: charles.hoge@us.army.mil.

© 2008 International Society for Traumatic Stress Studies. Published online in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/jts.20341

“stressful military experience” in the PCL-M, an identified “stressful experience” in the PCL-S, and “stressful experiences” in general in the PCL-C. Respondents rate each item from 1 (*not at all*) to 5 (*extremely*) to indicate the degree to which they have been bothered by that particular symptom over the past month, i.e., current symptoms. Different scoring schemes are possible. The most commonly used method of scoring involves summing the responses from the 17 PCL items yielding a score range from 17 to 85 and selecting a cutoff for caseness within this range. Another scoring strategy for identifying PTSD relies on the *DSM-IV* symptom criteria (Posttraumatic Stress Disorder Checklist-Military [PCL-M], 2005). Subjects who report at least one reexperiencing symptom, three avoidance symptoms, and two hyperarousal symptoms at the moderate or higher level of distress on the PCL are considered positive for PTSD. The two strategies may be combined (Hoge et al., 2004; Ruggiero, Del Ben, Scotti, & Rabalais, 2003).

The properties of the PCL were first presented by Frank Weathers and his colleagues in 1993. These included measures of test-retest reliability, internal consistency, convergent validity, and diagnostic utility. The optimal cutoff score for making the diagnosis of PTSD among male combat veterans was reported to be 50 (Weathers et al., 1993). More recent studies have reported that the optimal cutoff for identifying PTSD in various populations ranges from 30 to 50, and there are no clear guidelines to help users distinguish between the cutoffs recommended in these studies (Kang, Natelson, Mahan, Lee, & Murphy, 2003; Lang, Laffaye, Satz, Dresselhaus, & Stein, 2003; Sherman, Carlson, Wilson, Okeson, & McCubbin, 2005; Ventureyra, Yao, Cottraux, Note, & De Mey-Guillard, 2002; Walker, Newman, Dobie, Ciechanowski, & Katon, 2002). In addition to applications for screening, the PCL has been widely used to estimate the population prevalence of PTSD among military and veteran populations using the 50 point cutoff (Hoge et al., 2004; Kang et al., 2003; Smith, Smith, Jacobson, Corbeil, & Ryan, 2007), though it has been unclear if this is the appropriate cutoff for estimating population prevalence. Despite the widespread use of the PCL, there are no studies that assess the performance of the PCL on a population level or provide guidance regarding PCL cutoff selection in different populations.

The purpose of this study is to characterize the PCL's population-level performance extrapolating from published validation studies where the PCL was used in clinical care or other settings. We calculated the weighted average sensitivity and specificity values for PCL cutoff scores most widely reported across published studies where the PCL was compared with a gold-standard structured diagnostic interview. This information was then extrapolated to hypothetical populations with a known PTSD prevalence to demonstrate how the PCL would perform on a population level for screening or research purposes. The goal was not to identify an optimal efficiency or balance between sensitivity and specificity, as is typically done in validation studies, but rather to demonstrate the likely performance characteristics on a population level based on existing published sensitivity and specificity calculations.

This study provides valuable data regarding PCL's performance across published studies and its utility in population screening and prevalence estimation in epidemiological research. The study also is important in providing an example of how a commonly used PTSD screening measure, such as the PCL, is likely to perform on a population level based on its published test properties.

METHOD

Data Sources

Using the online database, PubMed (<http://www.ncbi.nlm.nih.gov/PubMed/>), a search was performed using the search string “PTSD checklist” during November 2007. The search produced 357 citations. Limiting the search to English language studies of humans reduced the number of citations to 328. The abstracts of all 328 citations were reviewed manually by one of the authors (AT) and those that mentioned the use of the PCL and any form of a structured clinical interview (gold standard) for the determination of the true prevalence of PTSD in the study sample were selected for further examination. More than 270 articles that did not involve the PCL and a gold standard were excluded in this manner. The remaining articles were examined to identify those that directly compared the PCL against structured clinical interviews: the Clinician Administered PTSD Scale (CAPS; Weathers, Ruscio, & Keane, 1999), the Structured Clinical Interview for DSM-IV (SCID; First, Spitzer, Gibbon, & Williams, 1996), or the Composite International Diagnostic Interview (CIDI; WHO, 1993; Wittchen, 1994) in different populations of U.S. or European adults. Fourteen articles met these criteria and were chosen for data abstraction: Andrykowski, Cordova, Studts, & Miller, 1998; Blanchard, Jones-Alexander, Buckley, & Forneris, 1996; Dobie et al., 2002; Forbes, Creamer, & Biddle, 2001; Grubaugh, Elhai, Cusack, Wells, & Frueh, 2007; Lang et al., 2003; Lang & Stein, 2005; Manne, Du Hamel, Gallelli, Sorgen, & Redd, 1998; Mueser et al., 2001; Sherman et al., 2005; Stein, McQuaid, Pedrelli, Lenox, & McCahill, 2000; Ventureyra et al., 2002; Walker et al., 2002; Yeager, Magruder, Knapp, Nicholas, & Frueh, 2007.

To verify the inclusion of the pertinent literature, an additional search was performed using the more specific PILOTS database of the National Center for PTSD (<http://www.ncptsd.va.gov/>) using the search query “PTSD checklist or PCL” combined with “Clinician Administered PTSD Scale or CAPS,” “Structured Clinical Interview for DSM or SCID,” or “Composite International Diagnostic Interview or CIDI,” resulting in the selection of 102, 73, and 13 citations respectively. Combining the lists produced 150 distinct citations, including journals, chapters, and dissertations, the abstracts of which were all manually reviewed as described above resulting in the identification of six of the articles already identified through PubMed, but also five additional publications: Christopher, 2001; McDevitt-Murphy, Weathers, & Adkins, 2005; McDevitt-Murphy, Weathers, Adkins, & Daniels,

2005; Prins & Ouimette, 2004; Prins et al., 2004; Widows, Jacobsen, & Fields, 2000. The first report on the PCL by Weathers et al. (1993) was frequently cited in these studies and was therefore included in the analysis bringing the total studies selected to 20. Several parameters were abstracted from the selected articles to include first author, year of publication, the number of study participants, participant gender, type of trauma, a description of the study population, the PTSD prevalence based on structured diagnostic interviews, the PCL cutoff(s) used, the sensitivity and specificity values at each cutoff, and the applied gold standard. Of the selected articles two were excluded from further analysis because they only used the *DSM* scoring criteria for the PCL (Mueser et al., 2001; Stein et al., 2000). The article by Forbes et al. (2001) was excluded due to a design that included PCL administration before and after treatment interventions. Administration of the CAPS to only a small segment of the study participants resulted in the exclusion of the article by Ventureyra et al. (2002). The dissertation by Christopher (2001) involving 16 subjects was rejected due to validation involving repeated PCL measurements and considerable selection bias in sampling. One of the reports by McDevitt-Murphy et al. (McDevitt-Murphy, Weathers, Adkins, & Daniels, 2005) was excluded as it duplicated findings from an earlier article (McDevitt-Murphy, Weathers, & Adkins, 2005) using the same sample. The remaining 14 studies included in the analysis are displayed in Table 1.

Data Analysis

Our examination of PCL's performance was guided by principles relevant to the review of screening and diagnostic tests (Blackman, 2001; Deeks, 2001; Irwig et al., 1994; ter Riet, Kessels, & Bachmann, 2001). To adjust for explicit test thresholds, PCL cutoffs were grouped into five categories that represented the cutoffs most commonly reported in the literature. The categories were 30 & 32, 38 & 40, 44 & 45, 48 & 50, and 55, 56 & 60. Where more than one cutoff value was reported in a particular category, the cutoff reported as having the highest efficiency was selected to represent a study. For example, if a study reported the sensitivity and specificity for both 30 and 32, the value with the highest efficiency was selected, so that this study was only represented once in this category. The weighted average sensitivity and specificity values for each cutoff category were then calculated using weights that were based on the total number of study participants. The weighted sensitivity and specificity averages were then applied to hypothetical populations varying in true PTSD prevalence. The prevalence estimates calculated in this manner were compared with the true prevalences in these populations.

RESULTS

Information about the articles included in this review is presented in Table 1. Nine of the 14 studies were published in the last

5 years. Together they encompassed 2,407 participants, 59% of whom were women ($n = 1,424$). A majority (75%, $n = 1,800$) of participants were patients from clinical care settings. Trauma type varied by study and was often unreported when participants were Veterans Affairs primary care patients, possibly indicating a presumption of combat trauma. Consistent with the clinical populations involved in most of the studies the current (i.e., last month) PTSD prevalence was often much higher than the 3.5% 12-month PTSD prevalence reported in the National Comorbidity Survey Replication (Kessler, Chiu, Demler, Merikangas, & Walters, 2005). An added factor resulting in the pooled prevalence of 20% may be the higher proportion of women in the clinical samples (Kessler et al., 2006; Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995). The studies included information on a wide range of PCL cutoffs. Validation was most often carried out using the CAPS (Weathers et al., 1999) as the gold standard (7 of 14 studies involving 70%, or 1,691, participants). We observed a wide range of sensitivity and specificity values reported for the PCL across different populations.

Examination of the pooled data (Table 2) showed that sensitivity values ranged widely within and across cutoff categories. The intracategory range narrowed at the cutoff extreme of 30 (.78 to 1.00). Specificity values also varied widely across and within categories with some convergence in specificity with increasing cutoff scores. As anticipated, sensitivity values decreased and specificity values increased with higher cutoffs. Despite the wide variability, the weighted sensitivity and specificity averages from the 14 studies showed a consistent linear relationship to increasing cutoff values, with weighted average sensitivity decreasing from .85 to .39 and specificity increasing from .73 to .97 across the five cutoff categories (Table 2).

Figure 1 applies the weighted average sensitivities and specificities from the published studies to a hypothetical population of 1,000 persons with a true PTSD prevalence of 15%, such as a population of soldiers undergoing screening for PTSD after returning from combat duty in Iraq. At a cutoff of 30 or 32, the weighted average sensitivity of .85 and specificity of .73 from the pooled studies would result in 128 true positive and 621 true negative tests. Of the 150 persons who truly have PTSD, all but 22 (85%) would be identified at this cutoff. The positive predictive value, however, would be only 36% (128 of 357). Of the 1,000 total individuals in the population, 357 (or 36%) would screen positive. If this cutoff was used to estimate the prevalence of PTSD in this population, then researchers would report that 36% of the population met criteria for PTSD, whereas the true prevalence is only 15%, grossly overestimating prevalence. If the instrument was being used to identify at-risk individuals for clinical referral, then this result would have implications in terms of the clinical resources needed. The remaining 2×2 tables demonstrate what happens as the cutoff is raised. A cutoff of 48 or 50 results in an estimated prevalence that most closely matches the true prevalence (14% vs. 15%), and positive predictive value is over 50% at this

Table 1. Summary of Reviewed Literature

Author	Year	Gender		Trauma type/Study Population	Prevalence	Cutoff	Sensitivity	Specificity	Validation
		N	%						
Andrykowski et al.	1998	82	100%	Breast cancer/ Cancer survivors	.06	30	1.00	.83	SCID
						35	.80	.87	
						40	.60	.93	
						45	.60	.99	
						50	.60	.99	
Blanchard et al.	1996	40	8%	MVA & sexual assault/Research volunteers	.45	44	.94	.86	CAPS
		3	92%			50	.78	.86	
Dobie et al.	2002	282	100%	Varied/VA primary care patients	.36	30	.85	.64	CAPS
		15	34%			38	.79	.79	
		29	66%			44	.68	.86	
						50	.58	.92	
						60	.41	.97	
Grubaugh et al.	2007	44	34%	Varied/Clinical patients with psychotic illness	.59	32	1.00	.17	CAPS
			66%			40	.96	.28	
						44	.85	.44	
						50	.69	.67	
						56	.65	.83	
Lang et al. #1	2005	154	48%	Not reported/VA primary care patients	.16	30	.96	.59	CIDI
		74	52%			50	.54	.94	
Lang et al. #2	2003	49	100%	Varied/VA primary care patients	.31	26	.94	.52	CIDI
						30	.78	.71	
						35	.67	.84	
						40	.61	.94	
						50	.39	.94	
Manne et al.	1998	65	100%	Child's cancer/ Mothers of cancer survivors	.06	40	1.00	.77	SCID
						45	.75	.82	
						50	.75	.89	

Continued

Table 1. Continued

Author	Year	Gender		Trauma type/Study Population	Prevalence	Cutoff	Sensitivity	Specificity	Validation
		N	%						
McDevitt-Murphy et al.	2005	57	M	6 11%	.25	44	.80	.93	CAPS
			F	51 81%					
Prins et al. ^a	2003	167	M	57 34%	.26	48	.84	.90	CAPS
			F	110 66%					
Sherman et al.	2005	141	M	18 13%	.23 ^b	38	.85	.90	SCID
			F	123 87%					
Walker et al.	2002	261	F	261 100%	.11	25	.93	.55	CAPS
Weathers et al.	1993	123	M	123 100%	.54	50	.82	.83	SCID
Widows et al.	2000	102	M	23 23%	.05	50	.20	.95	SCID
			F	79 77%					

Continued

Table 1. Continued

Author	Year	N	Gender		Trauma type/Study Population	Prevalence	Cutoff	Sensitivity	Specificity	Validation
			N	%						
Yeager et al.	2007	840	M	79%	Not reported/VA	.11	30	.83	.80	CAPS
			F	21%	primary care patients					
Summary	1993–2007	2407	M	41%	Varied/Largely	.20	25–60	.14 to 1.00	.17 to .99	CAPS; 70% CIDI; 9% SCID; 21%
			F	59%	clinical patients N = 1800; 75% ^c					

Note. M = Male; F = female; MVA = motor vehicle accident; VA = Veterans Affairs; SCID = Structured Clinical Interview for DSM IV; CAPS = Clinician Administered PTSD Scale; CIDI = Composite International Diagnostic Interview; NCPTSD = National Center for PTSD.

^aIncludes two articles. ^bLifetime and current posttraumatic stress disorder. ^cIncludes the studies by the following first authors et al.: Dobie, Grubaugh, Lang (two articles), Prins, Sherman, Weathers, and Yeager.

level. However, at this high specificity cutoff, the false-negative rate is high; nearly half of individuals who truly have PTSD are not identified.

Figure 2 compares the estimated PTSD prevalence, calculated from the weighted sensitivity and specificity averages of each cutoff category, with the true PTSD prevalence for six different populations (A–F) with hypothetical true PTSD prevalences ranging from 5 to 55%. The results show, for example, that in population A where the true prevalence of PTSD is 5% (such as a general population sample) only a PCL cutoff of 55 or higher will result in an accurate prevalence estimate. Approximately 10% of this population is expected to screen positive for PTSD using a cutoff of 48 or 50. If a PCL cutoff of 30 is used, then the estimated prevalence will be nearly 30%, although the true prevalence is only 5%. In population B where the true prevalence of PTSD is 15%, PCL cutoff values of 48 or 50 will most accurately estimate the PTSD prevalence, but as demonstrated in Figure 1, this prevalence estimate is derived from a combination of true positives and false-positives. In populations where the true prevalence of PTSD is 30% or higher (such as specialty care settings), PCL cutoff values of over 40 will underestimate the true prevalence.

DISCUSSION

Posttraumatic stress disorder is a common condition in the general U.S. population, where it has an estimated lifetime prevalence of 6.8% (Kessler, Berglund, et al., 2005) and a 12-month prevalence of 3.5% (Kessler, Chiu, et al., 2005). PTSD prevalence is higher among combat veterans with the most commonly cited estimate of current PTSD being 15% (Dohrenwend et al., 2006; Kulka & Schlenger, 1990). Estimating PTSD prevalence on a population level using gold standards such as the CAPS is often not feasible because of the respondent burden, personnel, training, time, and resources required to conduct structured clinical interviews. Given this, self-report instruments, such as the PCL, offer great utility for estimating prevalence. In this study, we combined data from published validation studies of the PCL to calculate weighted average sensitivities and specificities across different cutoff categories and report on how the PCL is likely to function on a population level.

The PCL's sensitivity and specificity both varied widely, with specificity appearing to be more stable than sensitivity in terms of a narrower range of values within studies and across the cutoff categories (Table 1). The range of specificity values was wider in studies with a higher PTSD prevalence. Sensitivity varied widely, particularly in the studies of populations where the prevalence of PTSD was low, such as the study by Walker et al. (2002) involving a primary care sample with a PTSD prevalence of 11% or the study involving treated cancer patients by Widows et al. (2000) where the PTSD prevalence was only 5%.

A remarkable finding of this study was that despite the wide variability across published validation studies, the weighted

Table 2. Cutoff Categories and Weighted Sensitivity and Specificity

<i>N</i> Participants/ <i>n</i> Studies	Cutoff category	Sensitivity			Specificity		
		Sensitivity range	Weighted average	Sensitivity 95% CI	Specificity range	Weighted average	Specificity 95% CI
1712/7	30 & 32	.78–1.00	.85	.82–.89	.17–.83	.73	.65–.82
1764/8	38 & 40	.57–1.00	.72	.66–.79	.28–.94	.86	.81–.90
1812/9	44 & 45	.36–.94	.62	.53–.71	.44–.99	.90	.87–.94
2209/12	48 & 50	.20–.84	.54	.43–.64	.67–.99	.93	.90–.96
1427/4	55, 56, & 60	.14–.65	.39	.27–.51	.83–.98	.97	.96–.98

average PCL sensitivity and specificity showed an expected linear relationship over the range of cutoff values. The weighted sensitivity average from the pooled studies decreased from .85 to .39 as the weighted specificity average increased from .73 to .97 over a range of increasing PCL cutoffs between 30 and 60. The study showed that in populations with PTSD prevalence of 15% or lower, cutoff values below 44 are likely to substantially overestimate the prevalence of the disorder. Clinical rather than population objectives may alter the way one considers these findings. For example, if the PCL were used for clinical screening, cutoffs below 44 would likely result in a high rate of unnecessary referrals, but the majority of persons who truly have PTSD would be identified. In populations with high PTSD prevalence, such as in specialty mental health settings, cutoff values of 44 or higher will likely underestimate prevalence. One recent study compared the PCL with a brief structured interview using the MINI among combat soldiers returning from Iraq (Bliese et al., 2008). This study found that a cutoff value of 30–34 provided the optimal efficiency in predicting referral for clinical evaluation. The finding highlighted that when the PCL is being used as a clinical detection tool a low cutoff value may be preferable.

Without cutoff adjustments, the PCL is limited in its ability to accurately estimate population prevalence. This study demonstrates how the accuracy of a particular cutoff in estimating prevalence will vary depending on the true prevalence of the disorder in the population. Thus, for purposes of estimating population prevalence, cutoffs need to be calibrated based on the expected prevalence of the disorder in the population, a finding that is counterintuitive because it suggests that the only way to obtain an accurate estimate of population prevalence is to know the prevalence in advance. Prevalence information for such a calibration can be obtained by first applying a gold standard to a small representative sample of the population under study or potentially by relying on published studies from similar populations if these are available and considered to be representative. Although prevalence is important, the most important consideration in selecting a cutoff however is the purpose of the measure, with lower cutoff values that achieve greater sensitivity preferred when there is a need to minimize false-negatives, and higher specificity cutoff values cho-

sen to minimize false-positive results. If a study's goal is only to estimate prevalence in a general population sample, the predictions identified in this article would indicate that a highly specific cutoff criteria will yield the most accurate estimates, the main risk being an underestimate if the true prevalence is higher than expected. For clinical screening or estimating prevalence in clinical populations (for example, to project resource needs in specialty mental health clinics), a more sensitive cutoff is necessary.

Our study did not assess the sensitivity and specificity of applying the *DSM-IV* criteria to the PCL instead of using the total numeric score. Preliminary data suggest that *DSM* categorical scoring produces similar results as the higher specificity numeric cutoff criteria. One study (Widows et al., 2000) that compared the categorical *DSM-IV* PCL definition (at least one Criterion B, three Criterion C, and two Criterion D at the moderate-3 or higher level) with the SCID found *DSM-IV* categorical scoring to produce a sensitivity of .40 and a specificity of .97, similar to a cutoff score in the 50s. A different categorical *DSM* scoring system (requiring a score of 4 or greater on at least one Criterion B, two Criterion C, and one Criterion D) showed a sensitivity of .32 and a specificity of .94, also making this similar to a cutoff score in the 50s (Stein et al., 2000). In the population-level study by (Hoge et al., 2004), 18.0%–19.9% of 1,692 soldiers and marines surveyed with the PCL 3–4 months after return from Iraq met the *DSM-IV* categorical definition of PTSD. Our analysis of these data showed that using a cutoff of 44 would have achieved nearly identical prevalence figures (18.4%–19.9%; analysis not shown).

This study assessed the performance of the PCL when used on a population level for estimating the prevalence of PTSD. Our report does not include an analysis of PCL's efficiency as a clinical screening tool, but it is clear that there is a tradeoff between obtaining accurate population prevalence estimates and correctly identifying which individuals in the population have the disorder. The most important factors that determine how the PCL will perform on a population level are the cutoff used and the true prevalence of PTSD in the population. Factors that may improve the screening performance of the PCL include the addition of questions pertaining to trauma (Criterion A of PTSD) or questions assessing significant impairment (Criterion F of PTSD). Our results

Cutoff Category 30 & 32				Cutoff Category 38 & 40			
A	True		Total	B	True		Total
	PTSD (+)	PTSD (-)			PTSD (+)	PTSD (-)	
PCL Screen (+)	128	229	357	PCL Screen (+)	108	119	227
PCL Screen (-)	22	621	643	PCL Screen (-)	42	731	773
Total	150	850	1000	Total	150	850	1000

Cutoff Category 44 & 45				Cutoff Category 48 & 50			
C	True		Total	D	True		Total
	PTSD (+)	PTSD (-)			PTSD (+)	PTSD (-)	
PCL Screen (+)	93	85	178	PCL Screen (+)	81	59	140
PCL Screen (-)	57	765	822	PCL Screen (-)	69	791	860
Total	150	850	1000	Total	150	850	1000

Cutoff Category 55, 56 & 60			
E	True		Total
	PTSD (+)	PTSD (-)	
PCL Screen (+)	59	25	84
PCL Screen (-)	91	825	916
Total	150	850	1000

	A	B	C	D	E
Weighted average sensitivity	0.85	0.72	0.62	0.54	0.39
Weighted average specificity	0.73	0.86	0.90	0.93	0.97
False negatives	15%	28%	38%	46%	61%
Positive predictive value	36%	48%	52%	58%	70%
Positive screens	36%	23%	18%	14%	8%

Figure 1. Two × two tables showing performance of the PTSD Checklist at five different cutoffs in a hypothetical population of 1,000 persons with true prevalence of 15%. For example, at a cutoff of 30 & 32, as depicted in Frame A, 36% of the population ($n = 357$) would screen positive, despite the true prevalence of 15% ($n = 150$).

do not allow for conclusions regarding how PCL's performance in prevalence estimation may change with the addition of trauma or impairment questions. However, it is likely that such questions would improve the diagnostic accuracy of the instrument.

It should be noted that our findings may suffer from publication bias, the quality of the selected literature, and heterogeneity among the studies, including different gold standards of comparison. To improve the quality of the comparisons and reduce heterogeneity, we only examined articles with gold standards that were (a) ac-

ceptable instruments for gauging the true presence or absence of PTSD, and (b) applied to all patients, minimizing the chances for verification bias (Irwig et al., 1994). We could not summarize the literature with regard to many desired details such as PCL version used, CAPS scoring criteria, duration from trauma event to assessment, or independence of test observations due to the absence of this information in the primary studies or difficulty comparing these factors across studies. However, given the available data on the PCL, we are not aware of any other method to derive direct

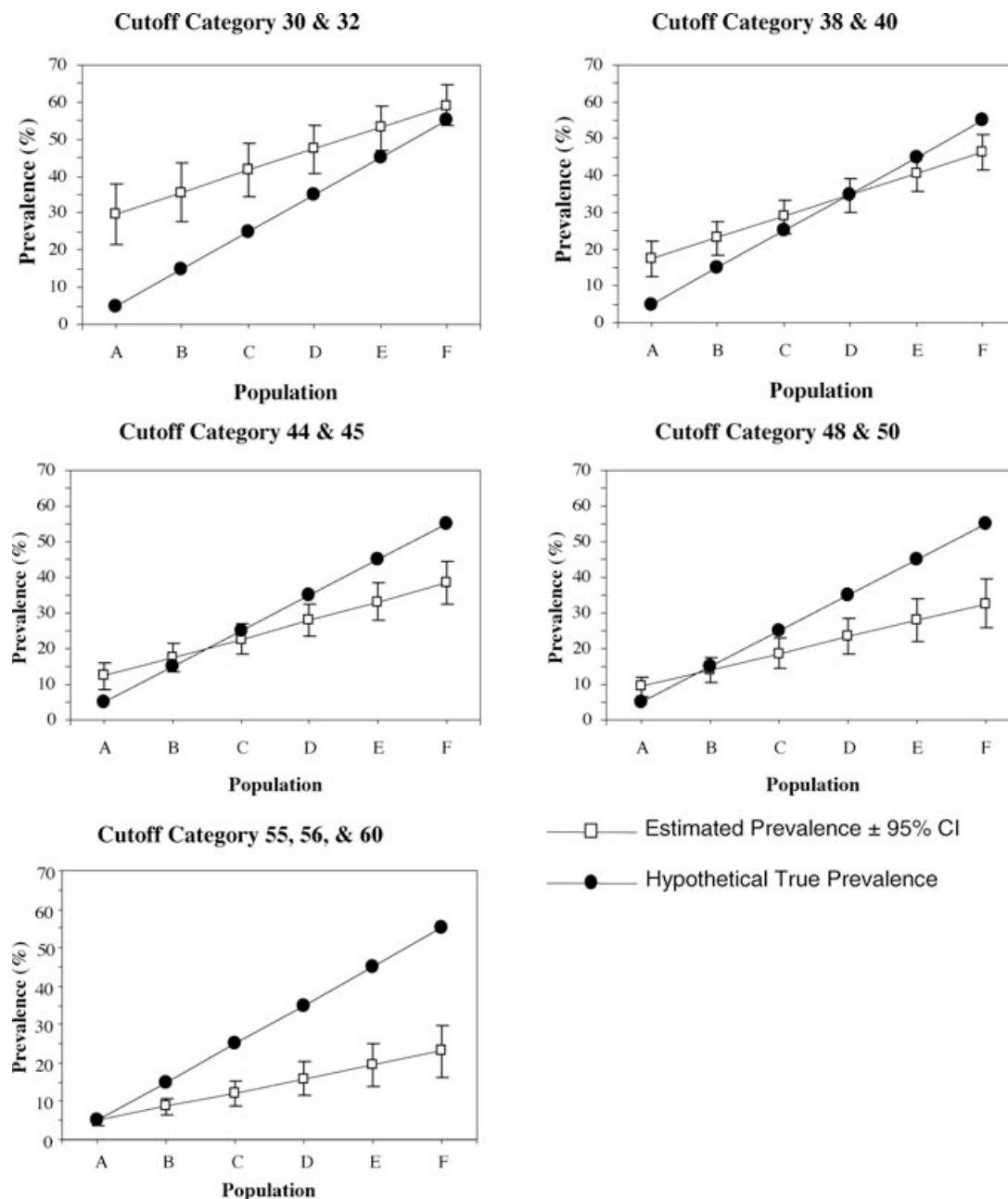


Figure 2. Comparison of estimated posttraumatic stress disorder (PTSD) prevalence to true prevalence for different PCL cutoff categories. Weighted sensitivity and specificity averages for each cutoff category were used to calculate the estimated prevalence in hypothetical populations A–F varying in true PTSD prevalence of 5, 15, 25, 35, 45, and 55%, respectively.

estimates of the PCL performance on a population level except by applying average test properties to populations with different prevalences. Indeed, instead of using weighted average sensitivities and specificities, we could have reached similar conclusions using hypothetical test values; for example, a test with an 80% sensitivity

and 80% specificity applied to a population with a 15% PTSD prevalence would result in a positive predictive value of 41% and a PTSD prevalence estimate of 29%.

Even the accepted gold standard measures show variability between them, and thus PCL performance would also be expected to

fluctuate between studies simply because of the different criterion measures applied and the fact that there is no definitive way to diagnosis PTSD with 100% accuracy. One study that evaluated nine different scoring criteria of the CAPS against the SCID in Vietnam veterans found that the criterion that had the strongest correspondence yielded a sensitivity of .91 and a specificity of .84 (Weathers et al., 1999). If this sensitivity and specificity were applied to a population with a 15% prevalence of PTSD, this would result in a positive predictive value of 50% and 27% of the population being identified as having PTSD, nearly double the true prevalence.

Our findings should be interpreted in light of the fact that the studies used for this review involved samples with a higher proportion of women. This may reflect the higher prevalence of PTSD among women and their greater relative tendency to seek clinical services (Bertakis, Azari, Helms, Callahan, & Robbins, 2000; Keene & Li, 2005; Mustard, Kaufert, Kozyskyj, & Mayer, 1998). To better assess the validity of the PCL among men, particularly active duty military personnel from the combat theaters of Iraq and Afghanistan, it is important that future validation studies be conducted in this population. The clinical setting of most studies and the nonrandom sampling of participants reduces the generalizability of our findings. However, the 14 studies encompass a pool of over 2,400 U.S. participants from a variety of settings, including primary care and specialty mental health clinics, and the aggregate results showed remarkably consistent linear trends in the direction expected for sensitivity and specificity values. These findings enhance the likelihood that the results may be generalizable. The study also highlights the limitations inherent in population use of this instrument as well as how clinical and population measurement objectives may conflict.

The methodology used in this study to extrapolate performance characteristics of a self-report tool with a known sensitivity and specificity on a population level are readily applicable to other instruments. For example, the 4-item Primary Care PTSD (PC-PTSD) screen (Prins & Ouimette, 2004; Prins et al., 2004) is being used routinely to screen all service members returning from deployment to Iraq and Afghanistan (Milliken, Auchterlonie, & Hoge, 2007), a policy that was implemented prior to any validation of the instrument in this population. Ideally the sensitivity and specificity should first be determined using a structured diagnostic interview gold standard in a sample of the population. However, in the absence of this, the published sensitivities and specificities from a primary care sample could have been extrapolated to predict how the test would perform on the population level and to estimate the impact of screening on health care resources. For example, extrapolating the published sensitivity of .91 and specificity of .72 for a PC-PTSD cutpoint of 2 to a population of soldiers with a true PTSD prevalence of 15% would result in identifying over 90% of those who actually had PTSD, but would also result in 37% of the entire population screening positive, a result similar to what would be observed if the PCL were utilized in this population at a 30 or 32 cutoff.

In conclusion, the PCL is an important self-report tool for clinical screening and estimation of current PTSD prevalence in various populations. Its performance is affected by the prevalence of PTSD in the population and selected cutoff. Cutoff calibration based on our summarized findings can enhance PCL's performance on a population level. Important factors may underline the variation in PCL's performance. Future research may allow for more comprehensive meta-analyses that assess the reasons for the variation in the reported performance measures of the PCL. The study methods are highly relevant for assessing the performance of any self-report measure of PTSD on a population level.

REFERENCES

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Andrykowski, M. A., Cordova, M. J., Studts, J. L., & Miller, T. W. (1998). Posttraumatic stress disorder after treatment for breast cancer: Prevalence of diagnosis and use of the PTSD Checklist-Civilian Version (PCL-C) as a screening instrument. *Journal of Consulting and Clinical Psychology*, 66, 586-590.
- Bertakis, K. D., Azari, R., Helms, L. J., Callahan, E. J., & Robbins, J. A. (2000). Gender differences in the utilization of health care services. *The Journal of Family Practice*, 49, 147-152.
- Blackman, N. J. (2001). Systematic reviews of evaluations of diagnostic and screening tests. Odds ratio is not independent of prevalence. *British Medical Journal*, 323, 1188.
- Blanchard, E. B., Jones-Alexander, J., Buckley, T. C., & Forneris, C. A. (1996). Psychometric properties of the PTSD Checklist (PCL). *Behavior Research and Therapy*, 34, 669-673.
- Bliese, P. D., Wright, K. M., Adler, A. B., Cabrera, O., Castro, C. A., & Hoge, C. W. (2008). Validating the PC-PTSD and the PTSD Checklist with soldiers returning from combat. *Journal of Consulting and Clinical Psychology*, 76, 272-281.
- Christopher, S. R. (2001). *Utility of the Posttraumatic Stress Disorder Checklist in parents of children with developmental disabilities*. Unpublished doctoral dissertation, University of Tulsa, Tulsa, OK.
- Deeks, J. J. (2001). Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *British Medical Journal*, 323, 157-162.
- Dobie, D. J., Kivlahan, D. R., Maynard, C., Bush, K. R., McFall, M., Epler, A. J., et al. (2002). Screening for post-traumatic stress disorder in female Veteran's Affairs patients: Validation of the PTSD checklist. *General Hospital Psychiatry*, 24, 367-374.
- Dohrenwend, B. P., Turner, J. B., Turse, N. A., Adams, B. G., Koenen, K. C., & Marshall, R. (2006). The psychological risks of Vietnam for U.S. veterans: A revisit with new data and methods. *Science*, 313, 979-982.
- Elhai, J. D., Gray, M. J., Kashdan, T. B., & Franklin, C. L. (2005). Which instruments are most commonly used to assess traumatic event exposure and posttraumatic effects?: A survey of traumatic stress professionals. *Journal of Traumatic Stress*, 18, 541-545.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (1996). *Structured Clinical Interview for the DSM-IV Axis I Disorders*.
- Forbes, D., Creamer, M., & Biddle, D. (2001). The validity of the PTSD checklist as a measure of symptomatic change in combat-related PTSD. *Behavior Research and Therapy*, 39, 977-986.
- Grubaugh, A. L., Elhai, J. D., Cusack, K. J., Wells, C., & Frueh, B. C. (2007). Screening for PTSD in public-sector mental health settings: The diagnostic utility of the PTSD checklist. *Depression and Anxiety*, 24, 124-129.

- Hoge, C. W., Castro, C. A., Messer, S. C., McGurk, D., Cotting, D. I., & Koffman, R. L. (2004). Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. *New England Journal of Medicine*, 351, 13–22.
- Irwig, L., Tosteson, A. N., Gatsonis, C., Lau, J., Colditz, G., Chalmers, T. C., et al. (1994). Guidelines for meta-analyses evaluating diagnostic tests. *Annals of Internal Medicine*, 120, 667–676.
- Kang, H. K., Natelson, B. H., Mahan, C. M., Lee, K. Y., & Murphy, F. M. (2003). Post-traumatic stress disorder and chronic fatigue syndrome-like illness among Gulf War veterans: A population-based survey of 30,000 veterans. *American Journal of Epidemiology*, 157, 141–148.
- Keene, J., & Li, X. (2005). Age and gender differences in health service utilization. *Journal of Public Health*, 27, 74–79.
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62, 593–602.
- Kessler, R. C., Chiu, W. T., Colpe, L., Demler, O., Merikangas, K. R., Walters, E. E., et al. (2006). The prevalence and correlates of serious mental illness in the National Comorbidity Survey Replication (NCS-R) (Document Number (SMA)-06-4195). Rockville, MD: Center for Substance Abuse and Mental Health Services Administration.
- Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62, 617–627.
- Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry*, 52, 1048–1060.
- Kulka, R. A., & Schlenger, W. E. (1990). *The National Vietnam Veterans Readjustment Study: Tables of findings and technical appendices*. New York: Brunner/Mazel.
- Lang, A. J., Laffaye, C., Satz, L. E., Dresselhaus, T. R., & Stein, M. B. (2003). Sensitivity and specificity of the PTSD checklist in detecting PTSD in female veterans in primary care. *Journal of Traumatic Stress*, 16, 257–264.
- Lang, A. J., & Stein, M. B. (2005). An abbreviated PTSD Checklist for use as a screening instrument in primary care. *Behavior Research and Therapy*, 43, 585–594.
- Manne, S. L., Du Hamel, K., Gallelli, K., Sorgen, K., & Redd, W. H. (1998). Posttraumatic stress disorder among mothers of pediatric cancer survivors: Diagnosis, comorbidity, and utility of the PTSD checklist as a screening instrument. *Journal of Pediatric Psychology*, 23, 357–366.
- McDevitt-Murphy, M. E., Weathers, F. W., & Adkins, J. W. (2005). The use of the trauma symptom inventory in the assessment of PTSD symptoms. *Journal of Traumatic Stress*, 18, 63–67.
- McDevitt-Murphy, M. E., Weathers, F. W., Adkins, J. W., & Daniels, J. B. (2005). Use of the Personality Assessment Inventory in assessment of posttraumatic stress disorder in women. *Journal of Psychopathology and Behavioral Assessment*, 27, 57–65.
- Milliken, C. S., Auchterlone, J. L., & Hoge, C. W. (2007). Longitudinal assessment of mental health problems among active and reserve component soldiers returning from the Iraq war. *Journal of the American Medical Association*, 298, 2141–2148.
- Mueser, K. T., Salyers, M. P., Rosenberg, S. D., Ford, J. D., Fox, L., & Carty, P. (2001). Psychometric evaluation of trauma and posttraumatic stress disorder assessments in persons with severe mental illness. *Psychological Assessment*, 13, 110–117.
- Mustard, C. A., Kaufert, P., Kozyrskyj, A., & Mayer, T. (1998). Sex differences in the use of health care services. *New England Journal of Medicine*, 338, 1678–1683.
- Posttraumatic Stress Disorder Checklist-Military (PCL-M). (2005). Retrieved May 25, 2007, from http://www.hsrd.research.va.gov/for_researchers/measurement/instrument/instrument_reviews2.cfm?detail=86
- Prins, A., & Ouimette, P. (2004). Corrigendum. *Primary Care Psychiatry*, 9, 151.
- Prins, A., Ouimette, P., Kimberling, R., Cameron, P. P., Hugelshofer, D. S., Shaw-Hegwer, J., et al. (2004). The primary care PTSD screen (PC-PTSD): Development and operating characteristics. *Primary Care Psychiatry*, 9, 9–14.
- Ruggiero, K. J., Del Ben, K., Scotti, J. R., & Rabalais, A. E. (2003). Psychometric properties of the PTSD Checklist-Civilian Version. *Journal of Traumatic Stress*, 16, 495–502.
- Sherman, J. J., Carlson, C. R., Wilson, J. F., Okeson, J. P., & McCubbin, J. A. (2005). Post-traumatic stress disorder among patients with orofacial pain. *Journal of Orofacial Pain*, 19, 309–317.
- Smith, T. C., Smith, B., Jacobson, I. G., Corbeil, T. E., & Ryan, M. A. (2007). Reliability of standard health assessment instruments in a large, population-based cohort study. *Annals of Epidemiology*, 17, 525–532.
- Stein, M. B., McQuaid, J. R., Pedrelli, P., Lenox, R., & McCahill, M. E. (2000). Posttraumatic stress disorder in the primary care medical setting. *General Hospital Psychiatry*, 22, 261–269.
- ter Riet, G., Kessels, A. G., & Bachmann, L. M. (2001). Systematic reviews of evaluations of diagnostic and screening tests. Two issues were simplified. *British Medical Journal*, 323, 1188.
- Ventureyra, V. A., Yao, S. N., Cottraux, J., Note, I., & De Mey-Guillard, C. (2002). The validation of the Posttraumatic Stress Disorder Checklist Scale in posttraumatic stress disorder and nonclinical subjects. *Psychotherapy and Psychosomatics*, 71, 47–53.
- Walker, E. A., Newman, E., Dobie, D. J., Ciechanowski, P., & Katon, W. (2002). Validation of the PTSD checklist in an HMO sample of women. *General Hospital Psychiatry*, 24, 375–380.
- Weathers, F. W., Ruscio, A. M., & Keane, T. M. (1999). Psychometric properties of nine scoring rules for the clinician-administered posttraumatic stress disorder scale. *Psychological Assessment*, 11, 124–133.
- Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993, October). The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility. Paper presented at the annual convention of the International Society for Traumatic Stress Studies, San Antonio, TX.
- WHO. (1993). *Composite International Diagnostic Interview*. Geneva/New York: Author.
- Widows, M. R., Jacobsen, P. B., & Fields, K. K. (2000). Relation of psychological vulnerability factors to posttraumatic stress disorder symptomatology in bone marrow transplant recipients. *Psychosomatic Medicine*, 62, 873–882.
- Wittchen, H. U. (1994). Reliability and validity studies of the WHO—Composite International Diagnostic Interview (CIDI): A critical review. *Journal of Psychiatric Research*, 28, 57–84.
- Yeager, D. E., Magruder, K. M., Knapp, R. G., Nicholas, J. S., & Frueh, B. C. (2007). Performance characteristics of the Posttraumatic Stress Disorder Checklist and SPAN in Veterans Affairs primary care settings. *General Hospital Psychiatry*, 29, 294–301.